Using Mean Similarity Dendrograms to Evaluate Classifications

John Van Sickle
Dynamac International, Inc.
200 SW 35th St., Corvallis, OR 97331

MS 96037, Journal of Agricultural, Biological and Environmental Statistics
To appear, Dec. 1997 issue

Abstract

Mean similarity dendrograms are introduced as a new graphical tool for evaluating classifications, based on sample data from replicate objects within each of the proposed classes. The dendrograms compare the mean similarities between objects within the same class to the mean similarity between objects in different classes. They were designed to complement multidimensional scaling plots and permutation tests of class structure. The dendrograms offer a concise picture of the overall strength of a classification as well as the compactness and isolation of individual classes. Although broadly applicable, the dendrograms were motivated by a need for easily-communicated assessments of land classifications that are intended to serve as geographic frameworks for environmental research and management. The dendrograms and other similarity-based tools are applied to a single-factor classification of fish communities sampled along a 281-km section of the Willamette River in Oregon, USA. In a second example, the tools are used to evaluate a two-factor classification of fish communities sampled in wadeable streams of Oregon's Cascade Mountains and Willamette Valley. The dendrograms help to assess the relative classification strengths of the two factors, factor interactions, and an alternative classification derived from cluster analysis.

Key Words: ecoregions; fish communities; permutation tests; multidimensional scaling; ordination; cluster analysis; classification.

1. Introduction

Environmental scientists and managers have drawn from a broad spectrum of methodologies for developing classifications of ecosystems and their components. Methods for ecological land classification, for example, range from strictly numerical approaches such as cluster analysis, to qualitative, map-based "regionalizations", in which large areas are progressively subdivided using rules deduced from general ecological principles at landscape scales (Huang and Ferng 1990; Conquest et al. 1994). Relatively few tools are available, however, for evaluating the strength and utility of a proposed classification, based on multivariate measurements of objects within each class.

This paper presents a novel graphical display, the *mean similarity dendrogram*, for evaluating low-order classifications. The dendrograms were designed to complement existing nonparametric ordinations and hypothesis tests that employ a pairwise measure of similarity between the objects being classified (Clarke and Green 1988; Smith et al. 1990). Taken together, these graphical and testing methods directly assess the degree to which objects within the same class are more similar to each other than they are to objects in different classes. For classes that have been chosen independently of the similarity matrix, permutation methods provide a test of the hypothesis that mean within-class similarities exceed the mean similarity between classes. The new dendrogram complements such tests by plotting within and between-class mean similarities in a compact format. In addition, the dendrogram format helps illustrate whether individual classes are *compact* (all objects within a class are highly similar to each other) and at the same time *isolated* (objects within a class are dissimilar to objects in all other classes) (Cormack 1971; Gordon 1981).

Although similarity-based assessments of classification strength are applicable in any grouping context, they appear particularly useful for evaluating ecosystem and/or land classifications that are intended as frameworks for environmental management and research (Omernik 1987; Omernik and Griffith 1991; Hughes et al. 1994). The assessments can help to choose one from among competing classifications of ecosystem types (e.g., Naiman et al. 1992) or geographic areas (Omernik and Griffith 1991). In addition, researchers and managers often wish to use a particular land classification derived from one set of ecosystem attributes (such as soils and topography) to group a different set of attributes (such as biological communities). One might then wish to assess the classification's strength for the second set of attributes. Finally, similarity measures and analyses are often preferable for ecological data types, such as species abundance lists, that do not satisfy the distributional assumptions of the conventional multivariate methods (multivariate analysis of variance, discriminant analysis) used to compare groups (Clarke and Warwick 1994; Ludwig and Reynolds 1988).

In this paper, I apply similarity-based graphical and testing methods to regional classifications of riverine ecosystems. Geographic regions that were originally delineated using factors such as soils, climatic variables, and topographic relief are evaluated for their ability to classify fish communities sampled from streams within the regions (Hughes et al. 1987, 1994; Whittier et al. 1988; Omernik and Griffith 1991). The data sets consist of lists of fish species sampled at each of several replicate stream sites within each regional class.

Mean similarity dendrograms are first defined and developed in the context of an example data set having a single, externally-determined classification factor with four classes. Permutation tests and multidimensional scaling ordinations are briefly reviewed and applied to the same data. The single-factor example demonstrates the ability of mean similarity dendrograms to compare the compactness and isolation of individual classes and to complement the tests and ordinations.

Dendrograms, along with an ordination and hypothesis tests, are then presented for a second example data set in which fish communities are classified *a priori* by two factors. In the two-factor example, the dendrograms help assess factor interactions and the relative classification strengths of the two factors. An *a posteriori* classification of the same data is also developed through cluster analysis. Hypothesis test results of the *a priori* and *a posteriori* classifications cannot be compared because such tests are not valid for the latter case. The compact dendrogram format, however, offers a useful visual comparison of the two approaches.

2. Site Similarity, Multidimensional Scaling, and Permutation Tests

2.1 Data Sources for Single-Factor Example

Hughes and Gammon (1987) sampled fish assemblages in 1983 at 26 sites along a 281-km stretch of the mainstem Willamette River in Oregon (Figure 1). They divided the stretch into four contiguous sections, beginning at the river mouth, based on channel depth and mapped channel gradients: a freshwater tidal section (Section A, kilometer 0 to 43, containing 7 sites), a flat pool section (Section B, kilometer 43 to 84, with 4 sites), a section with low map gradient (Section C, kilometer 84 to 212, with 4 sites), and a shallow, upper section with higher map gradient (Section D, kilometer 212 to 301, with 4 sites). Hughes and Gammon hypothesized that these four stretches would effectively classify the river's fish communities into distinct groups. The 1983 sampling effort repeated a fish survey carried out in 1944 at 21 sites along the same river stretch (Dimick and Merryfield, 1945), and a similar survey was conducted yet again in 1992 at 19 sites (TetraTech, 1993). Only a small number of sites were common to all three surveys.

Here, I use data from the three surveys (1944, 1983, and 1992) to illustrate similarity-based tools for assessing classifications. The goal is to evaluate the strength, and consistency over time, of Hughes and Gammon's (1987) four river sections for classifying fish assemblages in the mainstem Willamette.

2.2 Site Similarity

Let s_{ij} be the similarity between sites i and j , with $s_{ij}=s_{ji}$. Numerous similarity measures and their properties have been explored for use in ecological settings (Green, 1980; Washington, 1984; Digby and Kempton, 1987; Boyle et al. 1990). In this paper, I employ a single, simple measure that is widely used for comparing biological communities. The measure reflects the list of species present at each site, but not their abundances.

The *Jaccard* similarity between sites i and j, based on species presence/absence, is given by $s_{ij}=C_{ij}/(C_{ij}+U_i+U_j)$, where C_{ij} is the number of species common to the two sites, and U_i , U_j are the numbers unique to each site (Digby and Kempton 1987). Jaccard similarity is the proportion of the total number of species at two sites that are shared by the sites, and it ranges between 0 (no species in common) and 1 (identical species lists at the two sites).

A set of all pairwise similarities for N sites can be conveniently displayed as a lower triangular matrix (e.g., Table 1), with $N(N-1)/2$ unique similarities, a total that does not include the diagonal self-similarities ($s_{ii}=1$). In a proposed classification, the N sites each belong to one of K classes, so that class k has n_k sites, $k=1,2...K$. With the similarity matrix rows and columns appropriately ordered, the matrix can be partitioned into rectangular blocks of between-class similarities and triangular blocks of within-class similarities (Table 1).

2.3 Multidimensional Scaling

Multidimensional scaling (MDS; Kruskal and Wish 1978; Digby and Kempton 1987; Johnson and Wichern 1988) can provide an effective ordination of similarities among the N sites (Clarke and Green 1988; Smith et al. 1990). To create the MDS ordination, similarities are first converted to dissimilarities (distances) d_{ij}. For the Jaccard measure, as well as other similarities spanning the range (0,1), $d_{ij}=(1-s_{ij})$ (Digby and Kempton 1987). MDS then locates the N sites on a two- or three-dimensional scatter plot such that plotted intersite distances represent the $N(N-1)/2$ dissimilarities as faithfully as possible. The MDS algorithm determines plotting locations by minimizing a "STRESS" measure of the differences between plotted and true dissimilarities (Kruskal and Wish 1978).

With the sites (points) on the plot coded by class membership, the MDS plot can depict the strength of a proposed classification (Figure 2; Clarke and Green, 1988; Smith et al., 1990). If the classification is strong, then sites within the same class will appear clustered together on an MDS plot (each class is compact), and plotted distances between the clusters will be relatively large (classes are isolated).

Smith et al. (1990) advocate the use of metric MDS, in which plotted distances reflect the numerical magnitudes of dissimilarities. Clarke and his colleagues (Clarke and Green 1988; Clarke 1993; Clarke and Warwick 1994) prefer the robust, nonmetric version of MDS, in which plotted intersite distances are determined by trying to preserve only the rank order of the original dissimilarities (Kruskal and Wish 1978; Johnson and Wichern 1988).

Two- or three-dimensional MDS plots are valuable exploratory tools, but they can misrepresent class compactness and between-class distances, if dissimilarities are strongly multidimensional. The reduction of a high-dimensional configuration of N points, as represented by the original dissimilarity matrix, to a low-dimensional representation usually involves substantial distortion of at least some dissimilarities (i.e., STRESS is high). Nonmetric MDS ordinations generally have lower STRESS than their metric counterparts (Kruskal and Wish 1978), but the nonmetric plot axes are unitless and plotted distances between sites can only be interpreted ordinally.

The two-dimensional nonmetric MDS ordinations of the Willamette River sites (Figure 2), computed separately for each sampling year, each have STRESS values in the "fair" to "poor" range, indicating that relative distances between plotted sites can be misleading (Johnson and Wichern, 1988; Clarke, 1993). Nevertheless, the plots indicate distinct grouping of the sites by river sections A to D in all three years, with the weakest group separation seen in 1944.

2.4 Permutation Tests

Permutation procedures can test whether between-class similarities are, on average, less than within-class similarities, for a specific *a priori* classification (Mielke et al. 1976, 1981; Clarke and Green 1988; Smith et al. 1990). The test is based on a null hypothesis of "no class structure" (Gordon 1981, 1987; Everitt 1993). Here I will only sketch the approach, using Smith et al.'s (1990) test statistic and notation. Edgington (1995), Manly (1991), and Good (1994) give general introductions to permutation testing.

Let $M=\bar{B}/\bar{W}$ be the ratio of the mean $\bar{B}$ of all between-class similarities (contained in rectangular blocks of Table 1), to a mean $\bar{W}$ of all within-class similarities (contained in triangular blocks of Table 1). If $\bar{B}_{j,k}$ is the mean of all similarities such that one site is in class j while the other is in class k (off-diagonal entries of Table 1 Inset), then $\bar{B}=\sum_{j,k>j} n_j n_k \bar{B}_{j,k} / \sum_{j,k>j} n_j n_k$. Similarly, $\bar{W}=\sum_{k} (n_k/N)\,\bar{W}_k$, where $\bar{W}_k$ is the mean of all similarities such that both sites are in class k (diagonal entries of Table 1 Inset). Mielke (1979) argues that the factor n_k/N, with its unequal weighting of individual similarities when class sizes n_k are unequal, yields a more efficient estimate of $\bar{W}$.

If the "no class structure" null hypothesis were true, then between- and within-class similarities should be approximately equal, yielding a value of M close to 1. The permutation distribution of M under the null hypothesis can be estimated by repeatedly permuting the class labels of the sites, while keeping class sizes n_k constant, and calculating M for each permutation. A small value of $M=M_{obs}$ observed for the classification being tested, relative to the null hypothesis permutation distribution, indicates that between-class similarities are indeed smaller, on average, than within-class similarities. Specifically, the P-value for the test is the probability of obtaining, by chance, an M from this distribution that is less than or equal to M_{obs} . That is, $P=\sum_i I[M_i \le M_{obs}]/N_p$, where I[·] is the (0,1) indicator and N_p is the total number of possible permutations.

In general, $N_p = N!/(n_1!n_2!...n_K!)$, with a somewhat smaller total if two or more classes have equal sizes (Smith et al. 1990). For cases having N_p<10000, I used Berry's (1982) algorithm to enumerate all N_p permutations. For larger N_p , I selected a random subset of size N_R=10000 from the set of all permutations (Jackson and Somers, 1989). The P-value is then estimated as the proportion of $(N_R + 1)$ M-values (the randomized subset, plus M_{obs}) that do not exceed M_{obs} (Manly, 1991; Edgington, 1995).

For the Willamette sites from 1992 (Table 1), $M_{obs} = \bar{B}/\bar{W}$ = 0.35/0.61 = 0.57, a value that was less than all 10000 randomized M_i generated by site class permutations. Hence, P is estimated to be <0.001, giving quite strong evidence that the river sections effectively separate 1992 fish assemblages into distinct groups.

Permutation tests of class structure can be formulated in several ways. For example, since the total of all similarities in a matrix is constant over all permutations, either $\bar{W}$ or $\bar{B}$ alone is an equivalent statistic to M (Smith et al. 1990). Sokal and Rohlf's (1995) permutation statistic is a Mantel-type correlation between the partitioned dissimilarity matrix and a same-sized matrix containing 0's, with 1's in positions corresponding to between-class dissimilarity. Clarke and Green (1988) use medians, rather than means, of between- and within-class dissimilarities to construct a permutation statistic that is consistent with their other rank-based similarity analyses. Finally, Gordon (1994) applies the Mann-Whitney U-statistic to the ranks of combined dissimilarities within one class and between its objects and those outside the class, in order to assess the genuineness of that class. Because Gordon's

(1994) classes are *a posteriori* clusters, the permuted dissimilarity matrix itself is not an appropriate null model. Instead, Gordon (1994) bases his randomization tests on null models for the underlying multivariate data from which the dissimilarities are derived.

3. Mean Similarity Dendrograms

Smith et al. (1990) suggest that permutation tests be accompanied by examination of the matrix of mean within and between-class similarities (e.g., Table 1 Inset), to assess whether the differences between $\bar{B}_{j,k}$ and $\bar{W}_k$ have any practical, as opposed to statistical, significance (see Rohm et al. 1987, for an example). In addition, a significant *M* statistic could result if only one of several classes was compact and isolated, while the others were diffuse and poorly separated. Where more than two classes are involved, the mean similarity matrix could suggest follow-up tests for specific pairs of classes (Smith et al., 1990). The following sections show how mean similarities can be plotted in a dendrogram format, providing a clear and succinct summary of various features of a mean similarity matrix. The plots are especially useful for comparing several matrices.

3.1 Dendrograms for a Single Factor

For a single classification factor, one can construct a dendrogram with branches for each class joined at a node plotted at $\bar{B}$ (Figure 3). Branches terminate at $\bar{W}_k$, giving branch lengths of $\bar{W}_k$-$\bar{B}$ (Figure 3). Ideally, classes will be about equally isolated and compact; that is, the dendrogram's branches will be long and of approximately equal length. Unlike MDS plots, the mean similarity dendrogram has an accurate scale of the original similarity units. It is clear from Figure 3, for example, that site pairs in different river sections in 1992 shared, on average, fewer than 40% of their fish species, while pairs within the same section shared between about 55% and 70% of their species.

The mean similarity dendrogram provides an operational definition for the strength of a classification. In "strong" classification, classes are generally isolated ($\bar{B}$ is small) and relatively compact (each $\bar{W}_k$ is large relative to $\bar{B}$). A strong classification thus has a small value of *M*. The Willamette river sections could be regarded as a strong classification of fish communities in 1992 (Figure 3), although the practical significance of the plotted differences between $\bar{B}$ and $\bar{W}_k$ must ultimately be assessed on ecological, rather than statistical, grounds.

With its focus on individual within-class means, the dendrogram format of Figure 3 seems best suited to assessing overall classification strength and to displaying heterogeneity of within-class similarities across classes. In a clear analogy to one-way ANOVA, a single class with high $\bar{W}_k$ and a small sample size could produce a significant M statistic, even if separation between the other classes was relatively weak.

By plotting only the overall $\bar{B}$, the dendrogram format does not display the separation between all pairs of classes. One could link the branch ends with overlapping sets of two-pronged dendrograms to show all possible $\bar{B}_{j,k}$, but such a plot would be difficult to untangle visually for more than a very few classes. However, the dendrogram can effectively display selected between-class means that are of particular interest.

For example, because fish habitats change gradually and continuously along the Willamette River, one would expect site pairs from immediately adjacent river sections to be more similar than sites separated by one or two sections. This is confirmed by plotting dendrogram nodes at the means $\bar{B}_{j,k}$ of adjacent sections (immediate subdiagonal of Table 1 Inset). In 1992, between-section mean similarities for adjacent sections were consistently greater than the overall between-section mean $\bar{B}$, but they were still less than mean similarities within the adjacent sections (Figure 4).

If one is particularly interested in displaying the separation between all pairs of classes, then an ordination of all sites (Figure 2) will likely be more effective than a dendrogram. An attractive option is to use the matrix of mean similarities itself as the starting point for an ordination of the classes (Digby and Gower, 1981). Using principal coordinates analysis and a suitable transformation from similarities to squared distances, the centroids of K classes can be plotted in at most (K-1) dimensions with no distortion in mean between-class squared distance (Digby and Gower, 1981; Gower, 1966; Digby and Kempton, 1987). Once the class centroids have been plotted, Digby and Gower (1981) show how individual sites can be added to the plot so as to minimize distortion in the distances between each site and each class centroid.

The compact shape of the mean similarity dendrogram also facilitates comparison of multiple classifications. All three survey years on the Willamette River show that river sections strongly classify fish assemblages, with the 1983 and 1992 surveys having nearly the same patterns of between-class and much greater within-class similarities (Figure 5). In the 1944 dendrogram $\bar{W}_D$ is actually less than $\bar{B}$, indicating that fish communities were as variable within the upstream section D as they were throughout all four river sections (cf. the 1944 MDS plot in Figure 2). In this case, class D is not compact, relative to $\bar{B}$, even though the classification as a whole is fairly strong (M=0.77, P<0.001).

Apart from section D, the branch lengths of the 1944 dendrogram are similar to those of 1983 and 1992. However, the entire 1944 dendrogram is shifted toward smaller Jaccard similarities, relative to 1983 and 1992. Such a systematic shift in $\bar{B}$ and $\bar{W}_k$ would not be reliably detected on MDS plots. The lower mean simiarities for 1944 are explained by the sparser list of species in that year at all sites, as compared with 1983 and 1992. The Jaccard measure is known to be positively correlated with the total species count (Jackson et al., 1989). The sparse species lists in 1944 were probably due to the less efficient fish sampling methods used in the 1944 survey (Hughes and Gammon 1987). A more informative comparison of the river sections and their classification strengths across all three years would thus require reconciling the different sampling methods in 1944 and would also employ an alternative similarity measure, one that is resistant to differences in total species counts (Jackson et al., 1989).

4. Similarity Analyses for Two Factors

4.1 Data Sources and Evaluation Objectives for Two-Factor Example

Fish communities were sampled in wadeable streams in the Willamette Valley and Cascade Mountains of Oregon, as part of the US Environmental Protection Agency's Environmental Monitoring and Assessment Program, in cooperation with Oregon State University (Herlihy et al., 1996). The 28 streams discussed here (Figure 6) were sampled in 1993 and were classified by ecoregion (C=Cascades, V=Willamette Valley), and also by size

(S=Small, M=Medium). The ecoregion land classification delineates large areas having similar geology, climate, land surface forms, potential natural vegetation, soils, and land use. It was designed as a geographic framework suited to a wide range of resource management issues (Omernik 1987; Omernik and Griffith 1991). The stream size classes correspond approximately to Strahler stream orders 1 (Small) and either orders 2 or 3 (Medium) (Herlihy et al., 1996). Each of the 28 streams was sampled at one site and contributed a single observation to the data set.

The ecoregion-by-size classification is truly factorial, in the sense that Small and Medium streams have the same definition for both ecoregions. The factorial class structure poses several classification questions, arising out of clear parallels with factorial analysis of variance (Clarke 1993; Clarke and Warwick 1994). Do either or both factors offer significant classifying strength? Which of the two factors, ecoregion or size, has the greater classification strength and can be considered the "primary" factor? Finally, is there evidence for an interaction, in the sense that the classification strength of one factor is markedly different for the different levels of the other factor (Clarke 1993)?

Unlike in earlier studies of fish fauna in these ecoregions (Hughes et al. 1987; Whittier et al. 1988), the 28 sampled streams spanned a full spectrum of human disturbance histories and included both native and nonnative taxa. In addition, a few Valley streams partially drain areas in the adjacent Oregon Coast Range ecoregion and might be expected to take on characteristics of streams in that ecoregion. For these reasons, the classification strengths of ecoregion and size might be masked by high within-class variability.

4.2 Interaction and Relative Factor Strengths

A nonmetric MDS plot shows a clear left-right separation of Valley(V) and Cascades(C) streams (Figure 7). The Small (S) versus Medium (M) separation is not so clear, particularly for Valley streams, and this separation fails to emerge any farther in a three-dimensional MDS scatterplot (STRESS=0.11, plot not shown). Mean Jaccard similarities between Cascade and Valley streams are near zero, regardless of stream size (Table 2).

Mean similarities (Table 2) for this 2x2 classification can be conveniently displayed by plotting two-branched dendrograms for one factor (Ecoregion or Size) separately within each level of the other factor. The procedure is then repeated with the factors interchanged, resulting in four simple dendrograms (Figure 8). The four dendrograms and their associated permutation test results (Table 3) show that Ecoregion is a stronger classifying factor than Size. Branches are longer, because between-class similarities are smaller, for Ecoregion classes (upper two dendrograms, Figure 8) than for Size classes (lower two dendrograms). The high mean similarity within the set of small Cascade streams (Table 2, Figure 8) is due to five of the six streams containing the same, single fish species.

Size appears to be a somewhat stronger classifying factor in the Cascades streams than in the Valley, as shown by longer branches in the Cascades dendrogram (Figure 8) and by a smaller *M* for Size in Cascade streams than in Valley streams (Table 3). This evidence for an interaction between the two factors is not as clear when comparing the classification strengths of Ecoregions for the two Size classes. Even with some evidence of interaction, the four one-way M statistics of Table 3 are all smaller than would have occurred by chance, so it may be reasonable to consider an overall measure and test of Ecoregion strength across both Size classes. Following Clarke (1993) and Edgington (1995), one could average the Ecoregion *M* statistics for Small and Medium streams (Table 3), and then compare this with average *M*-values generated by permuting Ecoregion class assignments separately within the two Size classes.

4.3 Hierarchical Dendrograms

If factor interactions are small enough so that it is reasonable to evaluate the overall strength of a single factor, then a hierarchical dendrogram provides a concise display of mean similarities for multiple classification factors. In Figure 9, Ecoregion, being the stronger of the two factors, is designated as the primary factor, with stream Size as secondary. The Ecoregion (primary factor) dendrogram is plotted at the values of $\overline{B}$ and $\overline{W}_k$ calculated by pooling the Size classes within each Ecoregion. Note that the use of these pooled statistics in an overall test of a single factor would not be appropriate if interactions were nonnegligible or class sizes were unequal. Dendrograms for the secondary factor (Size) are then added for each level of the primary factor, with secondary factor branch ends and nodes plotted at $\overline{W}_k$ and $\overline{B}$, respectively, calculated within each level of the primary factor (Figure 9). This hierarchical format is also helpful in assessing relative factor strength. If the weaker of two factors (Size) is instead plotted as the primary factor, and the stronger (Ecoregion) is plotted secondarily, then the treelike shape seen in Figure 9 is lost.

An alternative format for Figure 9 is obtained by plotting the node for the Size classes within Cascade streams at $\bar{W}_{Cascade}$ rather than at $\bar{B}$ for Size, and doing likewise for Valley streams. The primary node (for Ecoregion) can then be plotted at the mean of all the similarities, that is, at the $\bar{W}$ value for all sites placed in a single class. The resulting dendrogram shows the consistency and magnitude of progressive increases in mean within-class similarity as one uses zero, one, or two factors to classify the sites. This within-means format is not as visually complex as the Figure 9 format and is easier to read if one has more than two factors and/or multiple classes per factor.

4.3 Application to an optimal classification

The concise format of the hierarchical dendrogram recommends its use for comparing alternative multiway classifications. As an illustration, the *a priori* Ecoregion-by-Size classification can be compared to *a posterior*i groupings derived from a clustering algorithm. Such comparisons have been frequently, but only qualitatively, used to evaluate ecoregions and other land classifications from site-level data (e.g., McDonough and Barr 1977; Hughes and Gammon 1987; Hughes et al. 1987; Omernik and Griffith 1991).

I performed a cluster analysis on the 28 Valley and Cascade streams, based on the Jaccard similarity matrix. I used group-average linkages to join sites and clusters (Everitt 1993), for consistency with the general strategy of comparing mean between and within-class similarities. The resulting dendrogram was cut at the similarity level of 0.22 (dissimilarity = 0.78) to yield a class structure similar in dimensions to the Ecoregion-by-Size classification (Figure 10). That is, four site clusters were selected in a two-level hierarchy. Two outlier sites were reassigned to the four main clusters, based on close comparison of their species lists (Figure 10).

At the top level, the cluster analysis created two groups exactly matching the Valley and Cascades classes (apart from the outliers), giving convincing evidence for the primacy and optimality of the Ecoregion classification (Figure 10). Each secondary cluster, however, contained a mixture of Small and Medium streams (Figure 10). As one would expect from an optimal classification, hierarchical mean similarity dendrograms have roughly equal branch lengths within each primary and secondary cluster (Upper panel, Figure 11). Comparison of the dendrograms for cluster-derived and Ecoregion-by-Size classifications reveals the extent to which stream Size is a suboptimal second-level classifier (Figure 11). The optimal second-level clusters do indeed have lower between-class mean similarities and longer branches than do the Size classes within Ecoregions (Figure 9), but the differences between the two dendrograms may not be significant ecologically. In such a case, little would be gained by searching for a second-level factor that classifies fish communities more strongly than stream Size.

More generally, mean similarity dendrograms can help assess the high-level partitions generated by any numerical clustering method. Given that clustering methods may even generate classes for random data (Everitt 1993), assessing the strength of cluster-derived classes is especially important.

5. Discussion

Mean similarity dendrograms convey classification strengths through conceptually simple comparisons of within and between-class similarities. As a result, they may prove to be an attractive, nontechnical tool for evaluating environmentally-oriented land classifications. The concise dendrogram format allows visual comparison of several different classifications, such as those produced by different similarity measures or clustering algorithms. Unlike most ordinations, the dendrograms depict class separation and compactness directly in the original units of the chosen similarity measure.

The dendrogram format displays the relative compactness of individual classes by plotting within-class mean similarity for each class. These mean similarities are compared to the average of all between-class similarities to assess the overall strength of a class structure. However, the dendrogram cannot easily display mean distances between all pairs of classes, and ordinations based on either the full similarity matrix or the reduced matrix of mean similarities are likely to give a clearer picture of relative class separation. For *a priori* classifications, plotting positions of dendrogram nodes and branches are the components of statistics used in permutation tests of "No class structure".

The extension of dendrograms to more than two factors is straightforward and limited only by the complexity of the resulting plot. In addition, the plotting format for two or more factors need not be confined to a crossed design. For example, a Size-by-Ecoregion dendrogram would still have been helpful if the Size class definitions for Cascade streams did not match those for Valley streams.

No attempt has been made here to show standard errors or confidence intervals for similarities on the dendrograms. The lack of independence among pairwise similarities makes it unclear how to interpret estimates of their variance and is an underlying reason for the necessity of permutation-based inference. Smith et al. (1990) note how inflated error rates can result when this dependence is overlooked and methods such as t-tests are applied to grouped similarities.

In this paper, I chose to plot the means of within and between-class similarities, for consistency with Smith et al.'s (1990) permutation statistic. But the dendrograms could just as easily display medians, as used by Clarke and Green (1988) in their permutation tests. Dendrograms could also be used to depict statistics of dissimilarities or distances, rather than similarities. Although dissimilarities can always be transformed to similarities by subtraction from a constant (Gordon 1981; Digby and Kempton 1987), there are some dissimilarity measures, such as Euclidean distance, that are more readily interpreted in their original units. A dissimilarity dendrogram could accompany a permutation test of whether mean between-class dissimilarity significantly exceeded the within-class mean (Smith et al. 1990).

Acknowledgments

I thank Francie Faure for her assistance with graphics. Two anonymous referees offered numerous valuable comments on the manuscript. I thank Tony Olsen and Glenn Griffith for critical reviews of an earlier draft. Brooke Abbruzzese, Joan Baker, Alan Herlihy, Bob Hughes, Manuela Huso, Phil Kaufmann, Dixon Landers, Phil Larsen, and Jim Omernik all contributed insights. Alan Herlihy provided the Cascades/Willamette Valley fish data for the two-factor example. The research described in this article has been funded by the U.S. Environmental Protection Agency and is a contribution of the Pacific Northwest Research Program. This document has been prepared at the EPA National Health and Environmental Effects Research Laboratory in Corvallis, Oregon, through contracts No. 68-C4-0019 to ManTech Environmental Technology, Inc., and No. 68-C5-0005 to Dynamac Corporation. It has been subjected to the Agency's peer review and it has been approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

References

Berry, K.J. (1982), "Algorithm AS179. Enumeration of All Permutations of Multi-sets with Fixed Repetition Numbers", *Applied Statistics*, 31, 169-173.

Boyle, T.P., Smillie, G.M., Anderson, J.C., and Beeson, D.R. (1990), "A Sensitivity Analysis of Nine Diversity and Seven Similarity Indices", *Research Journal of the Water Pollution Control Federation*, 6, 749-762.

Clarke, K.R. (1993), "Nonparametric Multivariate Analyses of Changes in Community Structure", *Australian Journal of Ecology*, 18, 117-143.

Clarke, K.R. and Green, R.H. (1988), "Statistical Design and Analysis for a 'Biological Effects' Study", *Marine Ecology - Progress Series*, 46, 213-226.

Clarke, K.R. and Warwick, R.M. (1994), "Similarity-based Testing for Community Pattern: The Two-way Layout with No Replication", *Marine Biology*, 118, 167-176.

Conquest, L.L., Ralph, S.C., and Naiman, R.J. (1994), "Implementation of Large-Scale Stream Monitoring Efforts: Sampling Design and Data Analysis Issues", in *Biological Monitoring of Aquatic Systems*, eds. S.L. Loeb and A. Spacie, Boca Raton, FL: Lewis Publishers, pp. 69-90.

Cormack, R.M. (1971), "A Review of Classification", *Journal of the Royal Statistical Society*, Series A, 134, 321-367.

Digby, P.G.N. and Gower, J.C. (1981), "Ordination Between- and Within-Groups Applied to Soil Classification", in *Down-to-Earth Statistics: Solutions Looking for Geological Problems* , Syracuse University Geological Contributions, 8, 63-75.

Digby, P.G.N. and Kempton, R.A. (1987), *Multivariate Analysis of Ecological Communities*, London: Chapman and Hall.

Dimick, R.E. and Merryfield, F. (1945), "The Fishes of the Willamette River System in Relation to Pollution", *Engineering Experiment Station Bulletin Series*, 20, 7-55. Corvallis, OR: Oregon State University.

Edgington, E.S. (1995), *Randomization Tests* (3rd ed.), New York: Marcel Dekker.

Everitt, B.S. (1993), *Cluster Analysis* (3rd ed.), London: Edward Arnold.

Good, P. (1994), *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*, New York: Springer-Verlag.

Gordon, A.D. (1981), *Classification*, London: Chapman and Hall.

Gordon, A.D. (1987), "A Review of Hierarchical Classification", *Journal of the Royal Statistical Society*, Series A, 150, 119-137.

Gordon, A.D. (1994), "Identifying Genuine Clusters in a Classification", *Computational Statistics and Data Analysis*, 18, 561-581.

Gower, J.C. (1966), "Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis", *Biometrika*, 53, 325-338.

Green, R.H. (1980), "Multivariate Approaches in Ecology: The Assessment of Ecologic Similarity", *Annual Review of Ecology and Systematics*, 11, 1-14.

Herlihy, A.T., Kaufmann, P.R., Reynolds, L., Li, J. and Robison, G. (1996), "Developing Indicators of Ecological Condition in the Willamette Basin -- An Overview of the Oregon Prepilot Study for EPA's EMAP program", in *River Quality: Dynamics and Restoration*, eds. A. Lanen and D. Dunnett, New York: Lewis Publishers/CRC Press, pp. 275-282.

Huang, S.L. and Ferng, J.J. (1990), "Applied Land Classification for Surface Water Quality Management: I. Watershed Classification", *Journal of Environmental Management*, 3, 107-126.

Hughes, R.M. and Gammon, J.R. (1987), "Longitudinal Changes in Fish Assemblages and Water Quality in the Willamette River, Oregon", *Transactions of the American Fisheries Society*, 116, 196-209.

Hughes, R.M., Rexstad, E. and Bond, C.E. (1987), "The Relationship of Aquatic Ecoregions, River Basins and Physiographic Provinces to the Icthyogeographic Regions of Oregon", *Copeia*, 1987, 423-432.

Hughes, R.M., Heiskary, S.A., Matthews, W.J., and Yoder, C.O. (1994), "Use of Ecoregions for Biological Monitoring", in *Biological Monitoring of Aquatic Systems*, eds. S.L. Loeb and A. Spacie, Boca Raton, FL: Lewis Publishers, pp. 125-151.

Jackson, D.A. and Somers, K.M. (1989), "Are probability estimates from the randomization model of Mantel's test stable?", *Canadian Journal of Zoology*, 67, 766-769.

Jackson, D.A., Somers, K.M. and Harvey, H.H. (1989), "Similarity measures: measures of co-occurrence and association, or simply measures of occurrence?", *American Naturalist*, 133, 436-453.

Johnson, R.A. and Wichern, D.W. (1988), *Applied Multivariate Analysis* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Kruskal, J.B. and Wish, M. (1978), *Multidimensional Scaling*. Beverly Hills, CA: Sage Publications.

Ludwig, J.A. and Reynolds, J.F. (1988), *Statistical Ecology*, New York: John Wiley & Sons.

Manly, B.F.J. (1991), *Randomization and Monte Carlo Methods in Biology*, London: Chapman and Hall.

McDonough, T.A., and Barr, W.C. (1977), "An Analysis of Fish Associations in Tennessee and Cumberland Drainage Impoundments", *Proceedings of the Annual Conference of the Southeast Association of Fish and Wildlife Agencies*, 31, 555-563.

Mielke, P.W. Jr. (1979), "On the Asymptotic Nonnormality of Null Distributions of MRPP Statistics", *Communications in Statistical Theory and Methods*, A8, 1541-1550.

Mielke, P.W. Jr., Berry, K.J., and Johnson, E.S. (1976), "Multiresponse Permutation Procedures for *A Priori* Classifications", *Communications in Statistical Theory and Methods,* A5, 1409-1424.

Mielke, P.W. Jr., Berry, K.J., Brockwell, P.J., and Williams, J.S. (1981), "A Class of Nonparametric Procedures Based on Multiresponse Permutation Procedures", *Biometrika*, 68, 720-724.

Naiman, R.J., Lonzarich, D.G., Beechie, T.J., and Ralph, S.C. (1992), "Stream Classification and the Assessment of Conservation Potential", in *River Conservation and Management*, ed. P.J. Boon, Chichester, England: Wiley, pp. 93-123.

Omernik, J.M. (1987), "Ecoregions of the Coterminous United States", *Annals of the Association of American Geographers,* 77, 118-125.

Omernik, J.M and Griffith, G.E. (1991), "Ecological Regions Versus Hydrologic Units: Frameworks for Managing Water Quality", *Journal of Soil and Water Conservation*, 46, 334-340.

Rohm, C.M., Giese, J.W, and Bennett, C.C. (1987), "Evaluation of an Aquatic Ecoregion Classification of Streams in Arkansas", *Journal of Freshwater Ecology,* 4, 127-140.

Smith, E.P., Pontasch, K.W., and Cairns, J. Jr. (1990), "Community Similarity and the Analysis of Multispecies Environmental Data: A Unified Statistical Approach", *Water Research,* 24, 507-514.

Sokal, R.R. and Rohlf, F.J. (1995), *Biometry* (3rd ed.), San Francisco: Freeman.

TetraTech, Inc. (1993), "Willamette River Basin Water Quality Study: Ecological Systems Investigation Component Report", Report TC 8983-09, Oregon Department of Environmental Quality, Portland, OR.

Washington, H.G. (1984), "Diversity, Biotic, and Similarity Indices: A Review with Special Relevance to Aquatic Ecosystems", *Water Research,* 18, 653-694.

Whittier, T.R., Hughes, R.M., and Larsen, D.P. (1988), "Correspondence Between Ecoregions and Spatial Patterns in Stream Ecosystems in Oregon", *Canadian Journal of Fisheries and Aquatic Science*s, 45, 1264-1278.

Table 1. Jaccard similarities between fish assemblages sampled at Willamette River sites, 1992. Sites are identified by their distance (km) from the river mouth in Portland, OR, and are grouped by river section. Inset: Mean similarities between and within river sections.

SECTION	SITE																			
	2	1																		
	6	.55	1																	
	10	.50	.78	1																
A	14	.40	.67	.86	1															
	27	.40	.50	.62	.71	1														
	35	.27	.50	.62	.71	.71	1													
	40	.31	.50	.45	.50	.67	.50	1												
	47	.25	.45	.40	.44	.30	.44	.45	1											
B	63	.25	.33	.40	.44	.30	.44	.45	.56	1										
	77	.27	.43	.38	.42	.42	.42	.67	.50	.64	1									
	93	.25	.33	.40	.44	.30	.44	.45	.75	.75	.50	1								
	124	.14	.21	.25	.27	.17	.27	.21	.50	.36	.27	.50	1							
	150	.20	.27	.31	.33	.23	.33	.36	.42	.55	.40	.55	.64	1						
C	182	.27	.33	.38	.31	.21	.21	.33	.38	.38	.38	.50	.73	.62	1					
	206	.21	.29	.33	.25	.15	.25	.29	.45	.45	.33	.60	.70	.73	.82	1				
	232	.07	.21	.25	.27	.27	.40	.31	.25	.25	.27	.36	.33	.39	.36	.42	1			
	240	.11	.29	.25	.27	.27	.36	.47	.33	.33	.41	.33	.50	.53	.50	.47	.62	1		
D	283	.12	.25	.29	.21	.21	.31	.33	.29	.29	.29	.38	.46	.50	.57	.67	.58	.60	1	
	296	.19	.30	.26	.21	.21	.28	.37	.33	.40	.40	.33	.39	.50	.47	.53	.47	.67	.65	1
	SITE	2	6	10	14	27	35	40	47	63	77	93	124	150	182	206	232	240	283	296
	SECTION				A					B				C				D		

MEAN SIMILARITY

A	.56			
B	.39	.61		
C	.26	.44	.70	
D	.26	.32	.47	.60
	A	B	C	D

Table 2. Mean Jaccard similarities and number of sites, for fish assemblages in wadeable Willamette Valley and Western Cascade streams, 1993. S=Small, M=Medium, C=Cascades, V=Valley.

	# Sites	SC	MC	SV	MV
SC	6	.67			
MC	7	.23	.31		
SV	5	.11	.09	.25	
MV	10	.03	.06	.21	.28

Table 3. Test statistic $M = \bar{B} / \bar{W}$ and P-value, for "No class structure" tests of a secondary factor (Ecoregion or Size), separately within each level of a primary factor (Size or Ecoregion). P-values are not adjusted for multiple tests.

Primary Factor Level	Secondary Factor	M	P
Medium Size	Ecoregion	0.22	<0.001
Small Size	Ecoregion	0.23	0.004*
Cascades Ecoregion	Size	0.49	0.005*
Valley Ecoregion	Size	0.78	0.004*

* Exact values, from all possible permutations.

Figure 1. Willamette River Basin, Oregon. Circles denote sites on mainstem Willamette River sampled for fish in 1983. Horizontal lines delimit river sections A to D.

Figure 2. Nonmetric multidimensional scaling ordinations of sites in Willamette River sections A to D, based on Jaccard similarity between fish assemblages, for fish surveys from 1944, 1983, and 1992.

Figure 3. Mean similarity dendrogram for fish assemblages in Willamette River sections A to D, 1992. Dendrogram node plotted at overall mean between-section similarity and branches terminate at mean within-section similarity for each section.

Figure 4. As in Figure 2, except adjacent river sections are joined at their mean between-section similarity. Dashed line denotes overall mean between-section similarity.

Figure 5. As in Figure 2, for comparison of 1944, 1983 and 1992 fish assemblages. M statistic and P-values are for a test of "No class structure" in each year, and parentheses contain number of sampled sites.

Figure 6. Cascades and Willamette Valley Ecoregions of western Oregon, showing sampling locations for Small (triangles) and Medium (circles) streams.

Figure 7. Nonmetric multidimensional scaling ordination of fish assemblages at 28 stream sites in Willamette Valley (V) and Cascades (C) ecoregions in 1993. Stream size denoted by S (Small) or M (Medium). SC(5) denotes 5 Small Cascades streams having the same, single fish species.

Figure 8. Mean similarity dendrogram for fish in wadeable Oregon streams in 1993, classified by Ecoregion (V=Valley, C=Cascades) within Size classes and also by Size (S=Small, M=Medium) within Ecoregion.

Figure 9. Hierarchical mean similarity dendrogram for wadeable Oregon streams, with Ecoregion as the primary factor and Size as the secondary factor.

Figure 10. Dendrogram for group-average clustering of wadeable Oregon streams, with sites labeled by Ecoregion and Size class (see Figure 7 legend). Dendrogram is cut (dashed line) to create four clusters (C1 to C4). Outlier sites (1) and (2) are reassigned to C3 and C1, respectively.

Figure 11. Mean similarity dendrograms for cluster-derived and Ecoregion-by-Size classifications of wadeable Oregon streams.

Figure 1:

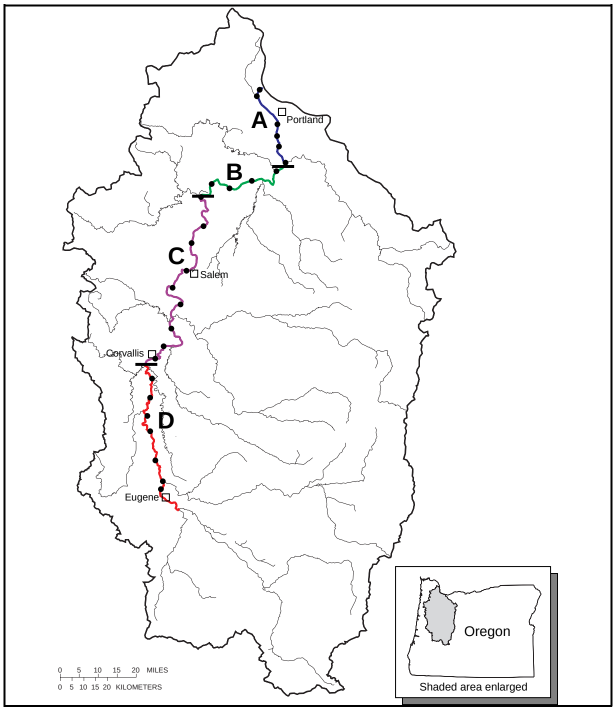

Figure 2

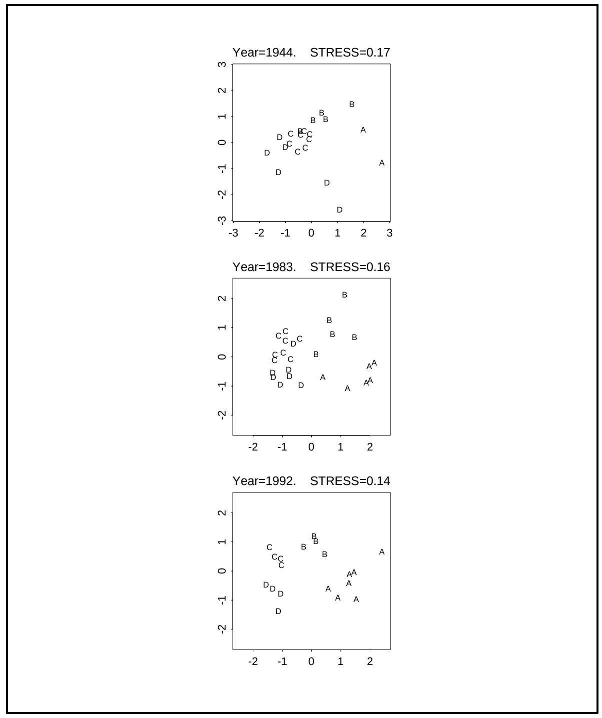

Figure 3:

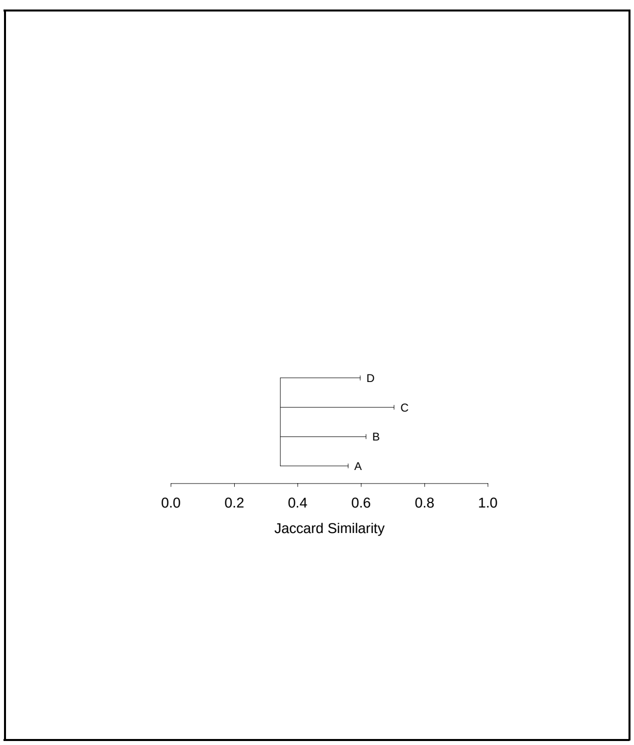

Figure 4:

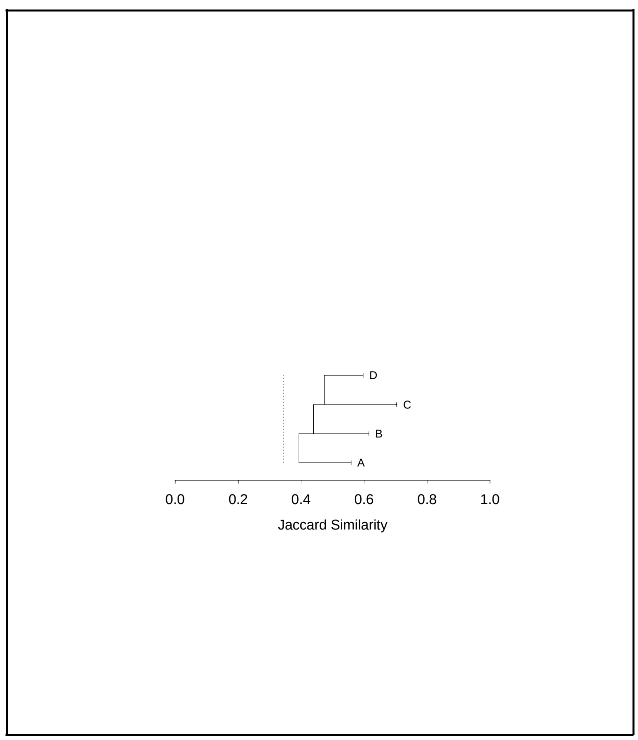

Figure 5:

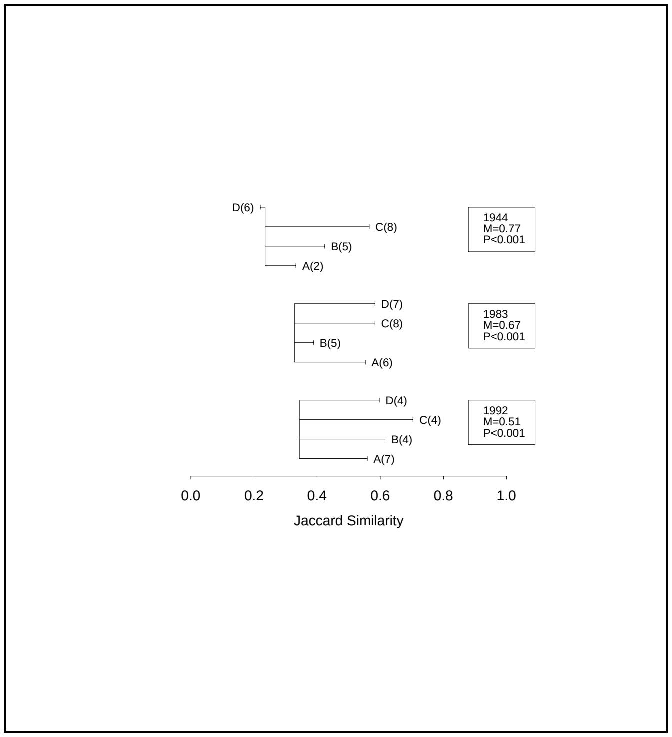

Figure 6:

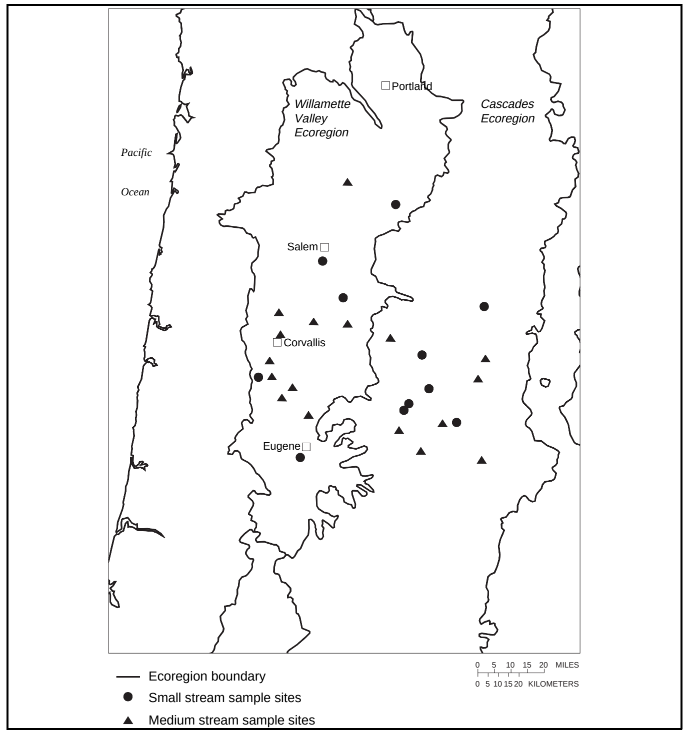

Figure 7:

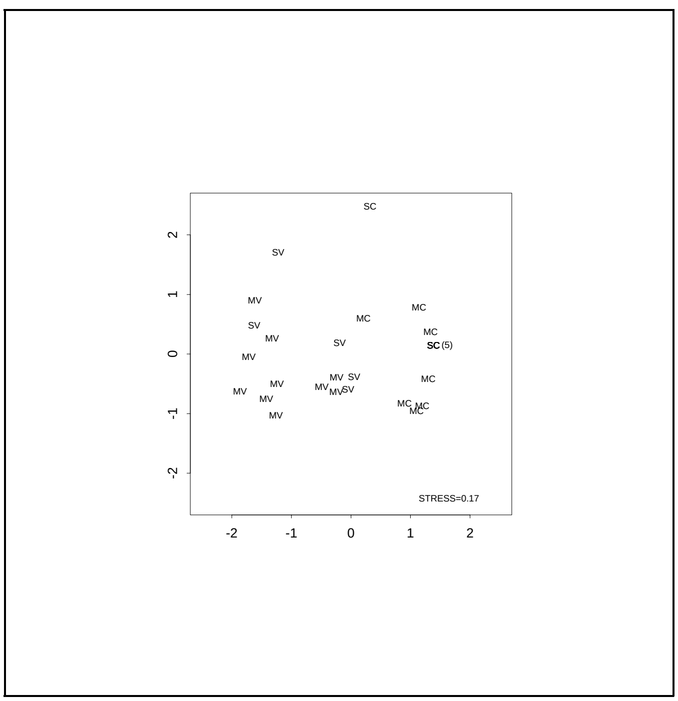

Figure 8

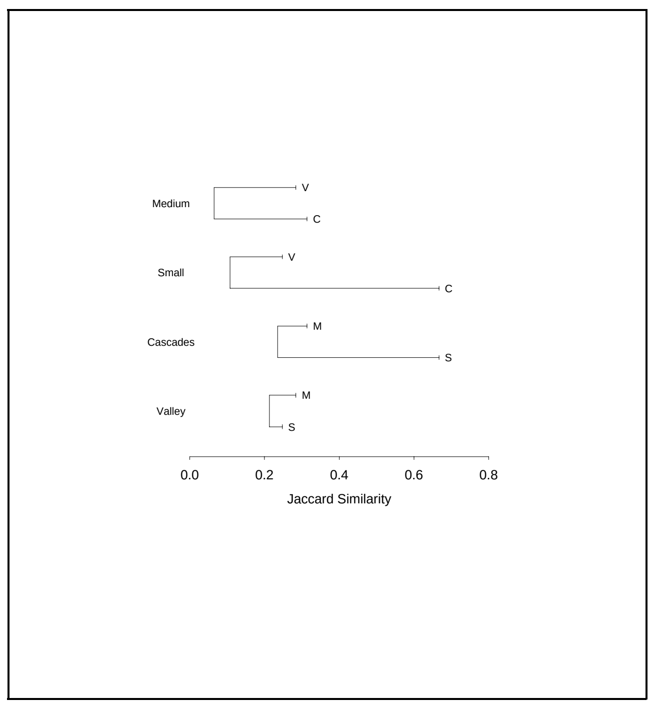

Figure 9:

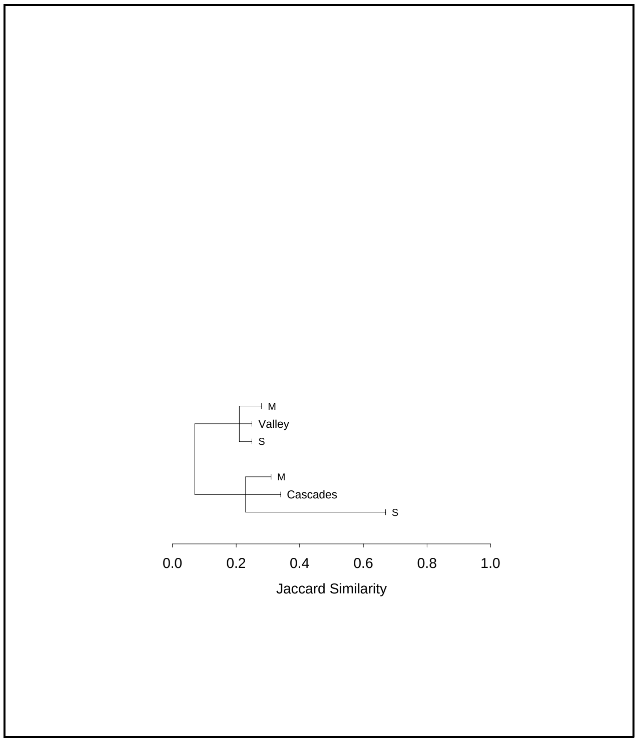

Figure 10:

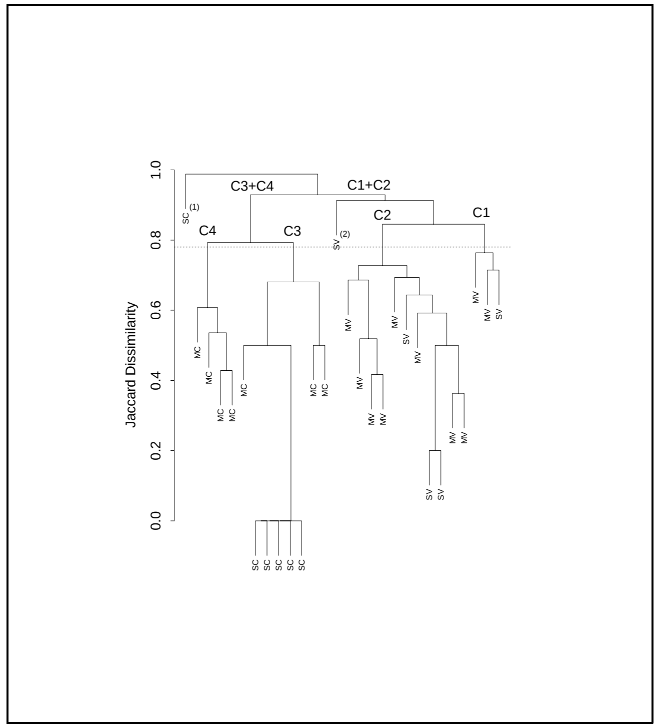

Figure 11: